Linkable Geographic Ontologies

Preview of GIR'10 paper

Francisco J. Lopez-Pellicer[1] Mário J. Silva[2] Marcirio Chaves[2]

[1]IAAA
Universidad de Zaragoza
Zargoza, Spain

[2]XLDB – LaSIGE
Universidade de Lisboa
Lisboa, Portugal

Grease-II Workshop
University of Lisbon, 26th January, 2010, 2:00PM, Room 6.3.35

Outline

Outline

Our problem

- Definition:

  The provision of large datasets of related geographic named places for geographic information retrieval tasks.

- Issues:

  1. Content : How to made their content more useful?
  2. Standards : How to share them?
  3. Integration : How to relate their content?
  4. Interface and access : How to ease the access and use?
  5. Other : How to convert from other sources?

Make them (RDF) linkable

- Our approach:

  A "small o" ontology of geographic names ready to be published as Linked Data.

- Summary:

  1. Content : Define a geographic RDF vocabulary and/or reuse RDF vocabularies (e.g. SIOC, FOAF, Basic Geo).
  2. Standards : RDF for data exchange.
  3. Integration : Use RDF links to RDF datasets (e.g. DBPedia) and non-RDF datasets (e.g. GNIS).
  4. Interface and access : Publish them as Linked Data and offer a SPARQL end-point.
  5. Other : Transform to RDF applying your content model.

"Small o" ontology

- A "big O" ontology deals with soundness and expressivity.
  - Model the World as expressively as possible.
  - Sound answers.
    - Each wrong answer costs money.
- A "small o" ontology deals with relations and people's tags
  - Model a world and apply to large and heterogeneous data collections.
  - Best effort answers.
    - Our systems can tolerate ambiguous anwsers.

Linked Data

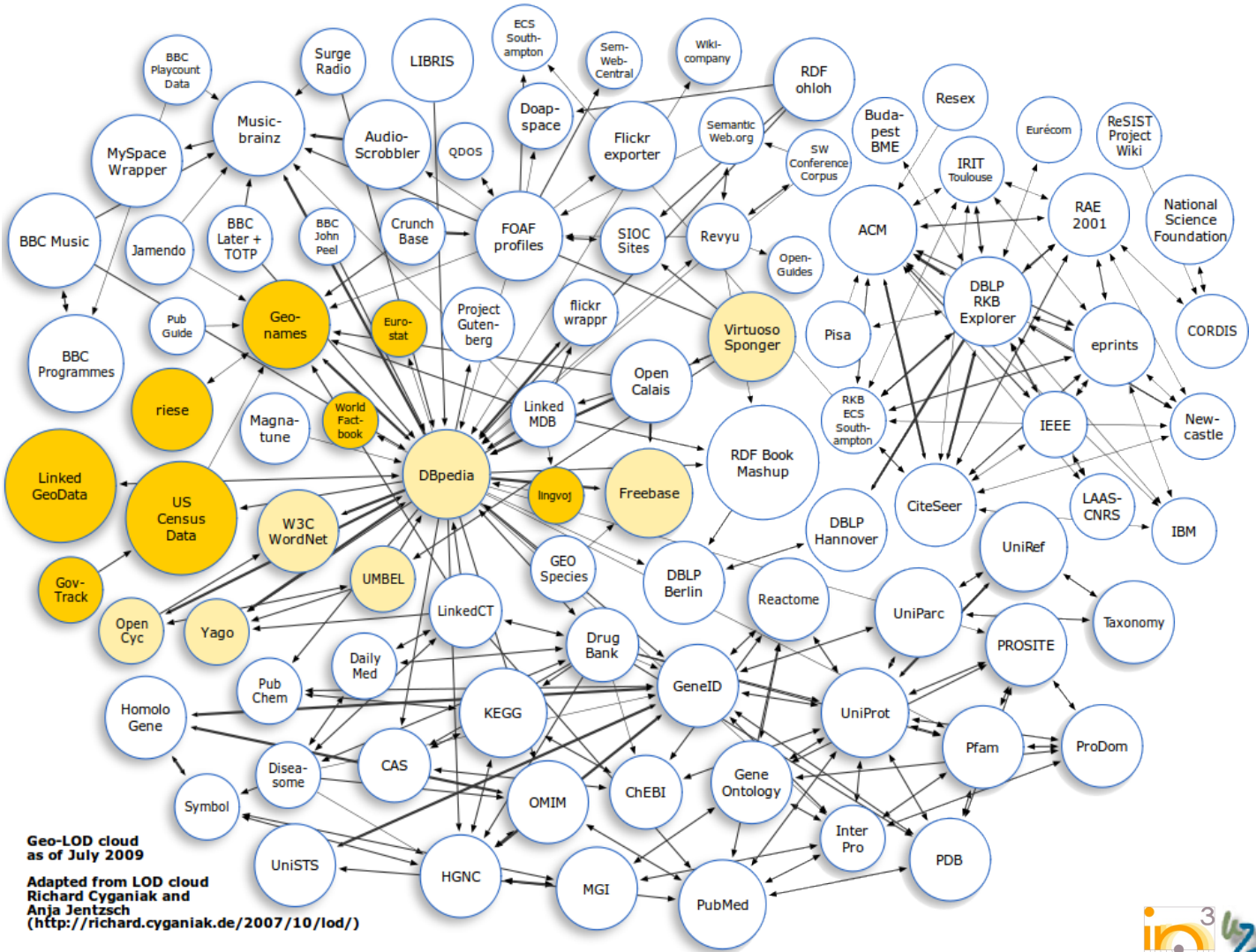
Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods.

`http://linkeddata.org/`

A term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF.

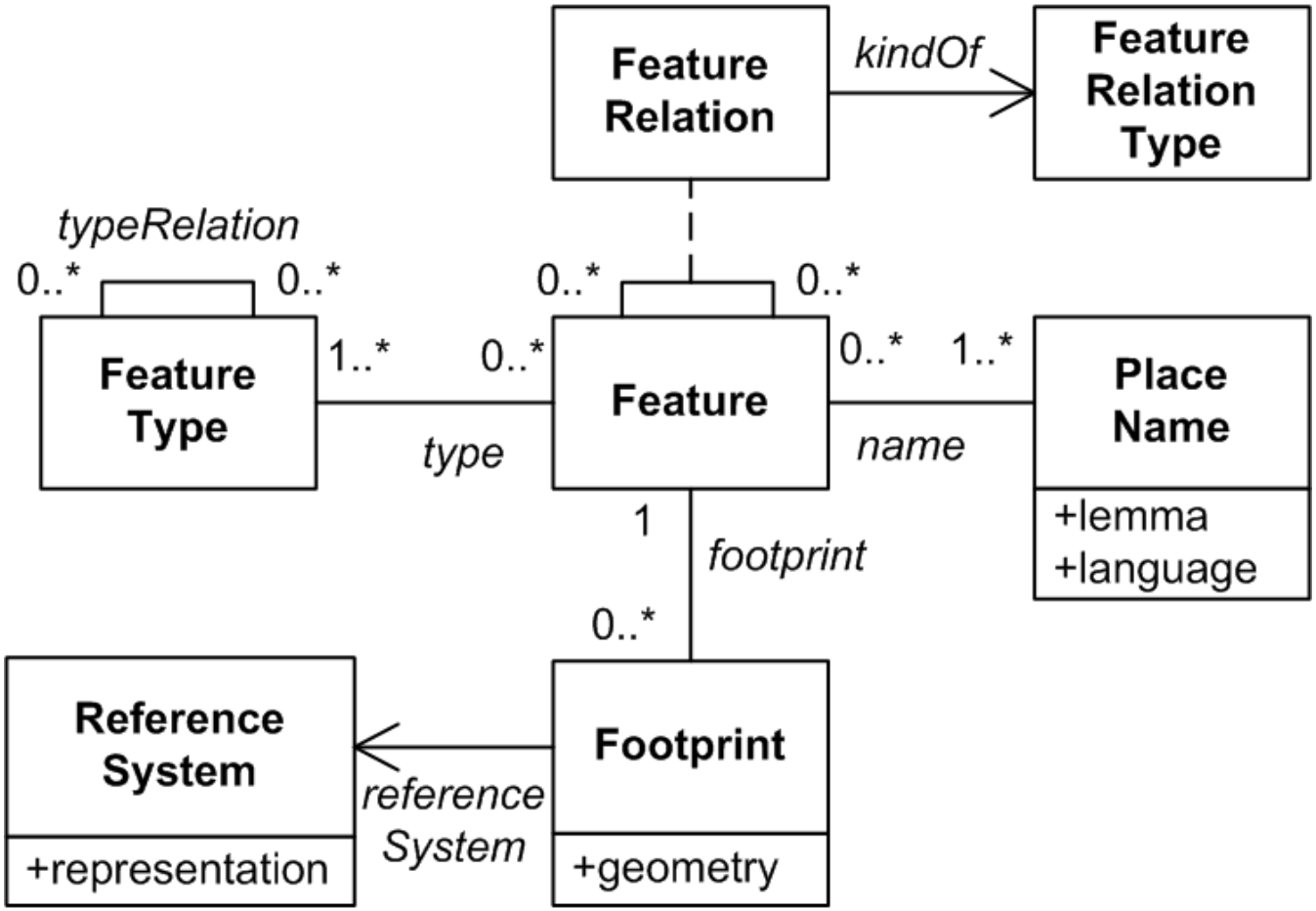
`http://en.wikipedia.org/wiki/Linked_Data`

Look to the LOD cloud



Outline

A world model



RDF Vocabulary

- Core
  - Classes Feature, PlaceName, FeatureType and Footprint.
  - Properties name, type, footprint, lemma, languageCode, geometry
- Modules
  - Information domains
  - Provenance
  - Spatial reference systems

How to use the vocabulary

- Describe the feature "Lisboa"

  Lisboa (38°42´N 9°11´W, Lisbon in English) is the capital of Portugal

- In RDF
  - Using TURTLE, a terse representation of RDF
  - With Dublin Core and Basic Geo annotations.

```
:Lisboa a gn:Feature ;
     dc:title "Lisboa"@pt , "Lisbon"@en ;
     dc:description "The capital of Portugal"@en ;
     rdfs:seeAlso <http://dbpedia.org/resource/Lisbon> ;
     geo:lat "38.7"^^xsd:double ;
     geo:long "-9.183333"^^xsd:double ;
     gn:name :LisboaName , :LisbonName ;
     gn:type :City ;
     gn:footprint :LisboaFootprint ;
     gn:relation :Portugal .
```

How to use the vocabulary (cont.)

- In RDF (cont.)
  - Provenance annotations.

```
:LisbonName a gn:Name ;
    gn:lemma "Lisbon"^^xsd:string ;
    gn:languageCode "en"^^xsd:string ;
    gn:lineage :Wikipedia .
```

  - Location data in GML.

```
:LisboaFootprint a gn:Footprint ;
    gn:geometry "<gml:Point xmlns:gml='http://www.opengis.net/gml'>
<gml:pos>38.7, -9.183333</gml:pos></gml:Point>"^^rdf:XMLLiteral ;
    gn:ReferenceSystem <urn:ogc:def:crs:EPSG::4326> ;
    gn:lineage :Wikipedia .
```

  - Example of application: geographic scope of a web page.

```
<http://www.cm-lisboa.pt/> a rdf:Resource ;
    dc:title "Webpage of the Lisbon City Council"@en ;
    gn:relation :Lisboa .
```

Outline

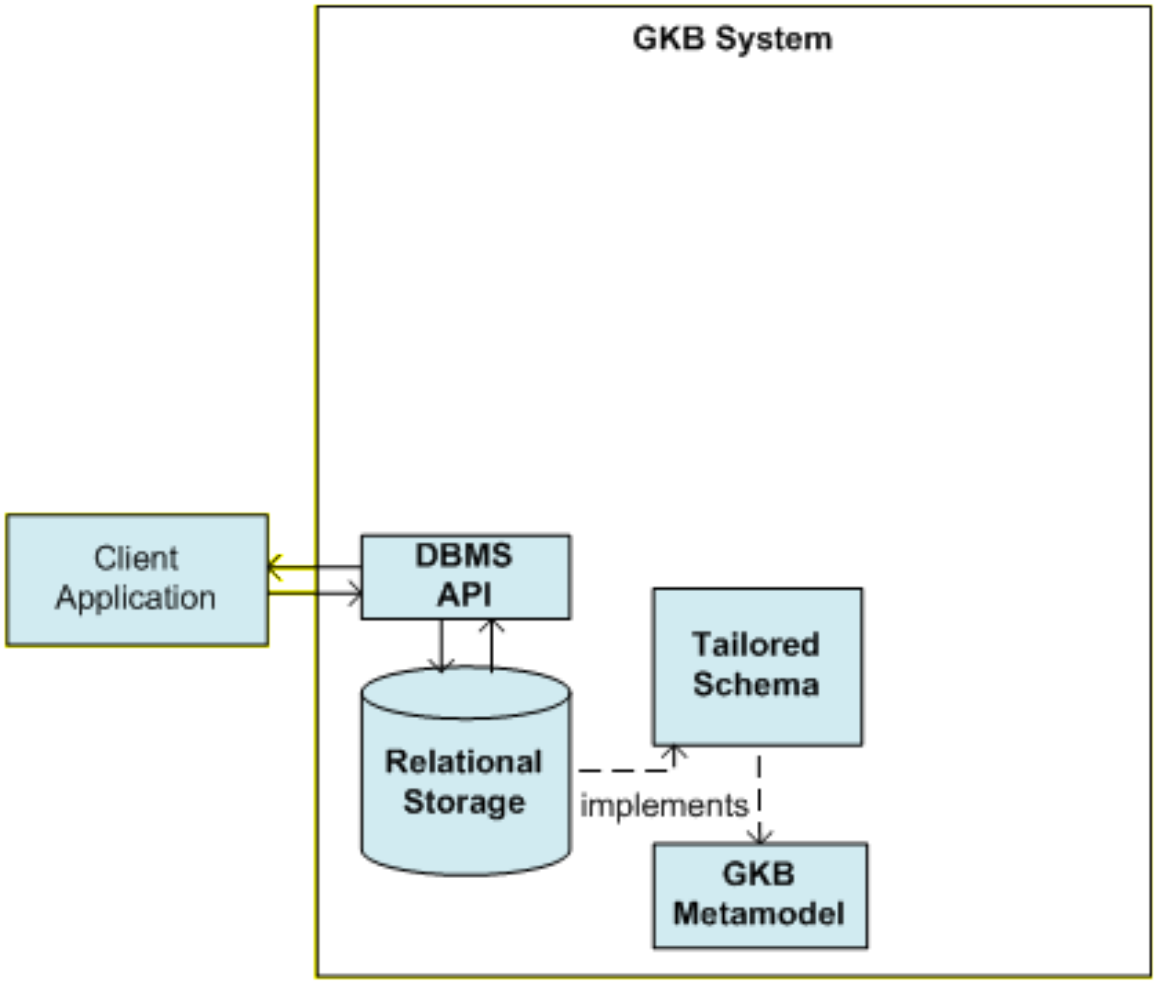
GKB system

- Geographic Knowledge Base (GKB) system

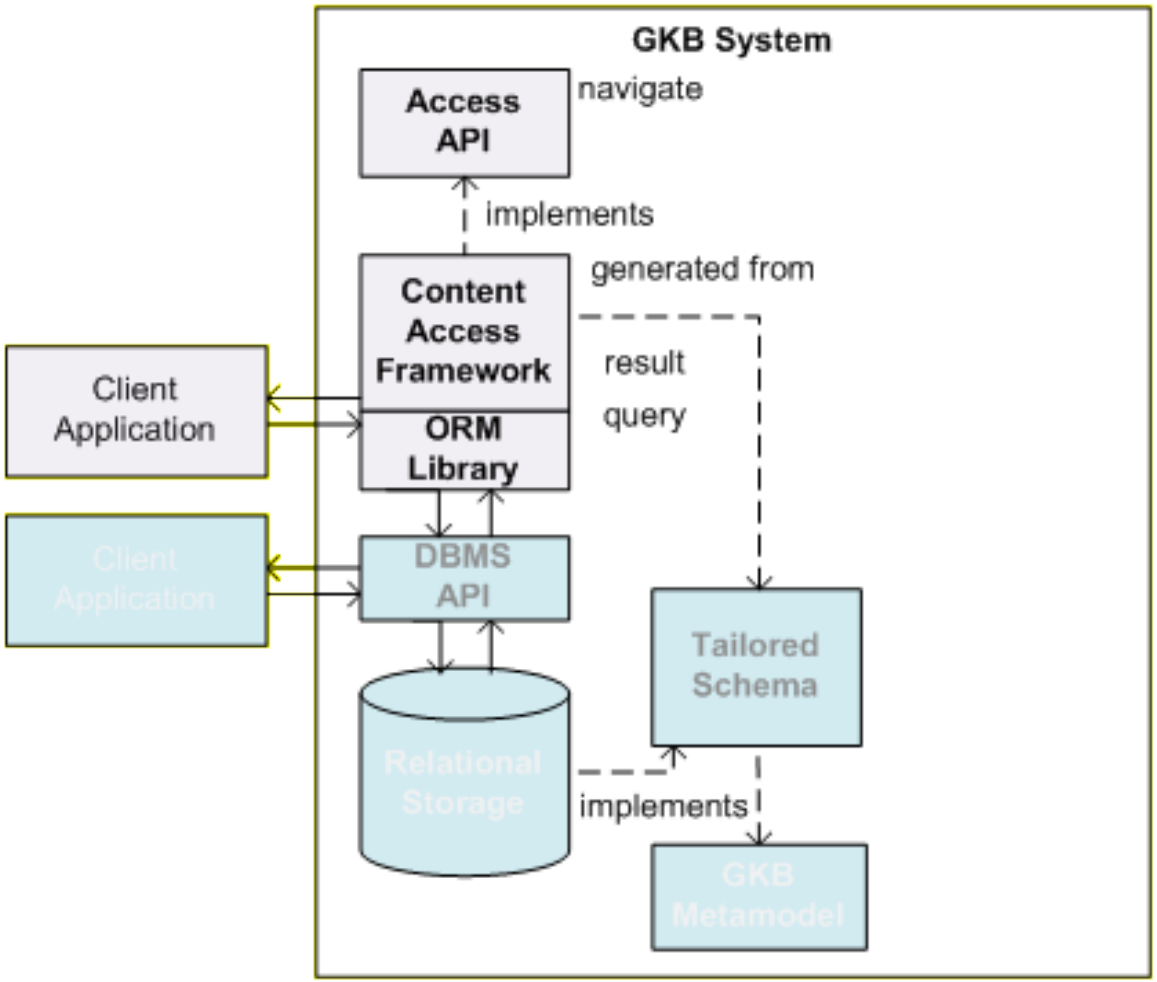
  A geographic knowledge management system that implements a domain-independent geographic meta-model (Geo-Net vocabulary) and integrates geographic knowledge collected from multiple sources.

- GKB is for:
  - Build and manage geographic ontologies.
  - Support IR tasks that require:
    - SQL queries
    - Object-oriented libraries
    - Semantic networks

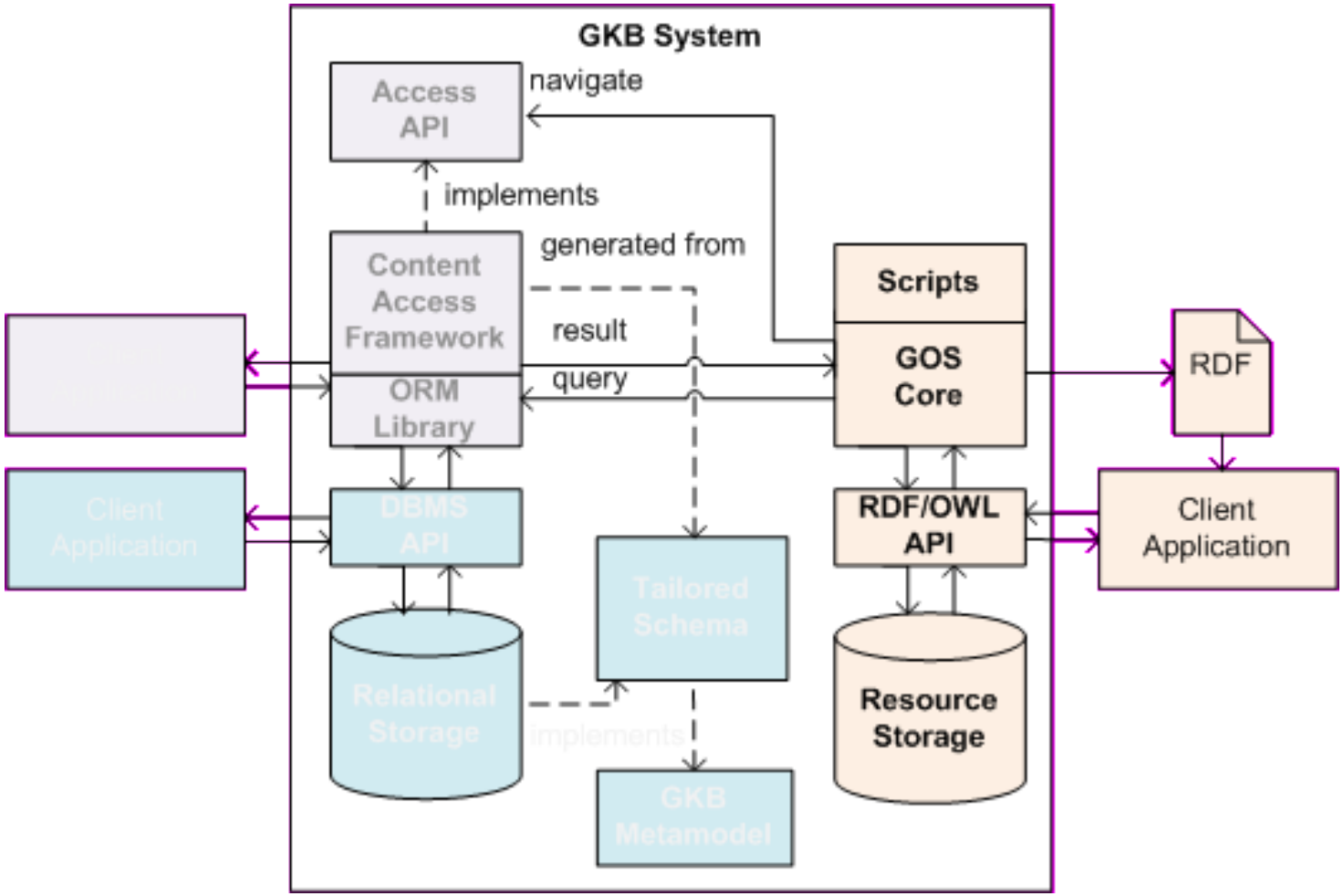
Relational persistence system



Object-oriented persistence system



Resource-oriented persistence system



Technical details

- Relational
  - PostgreSQL 8.3.6 + PostGIS 1.3.6
- Object-oriented
  - Java + Hibernate 3.3+ Hibernate Spatial / Java Topology Suite
- Resource-oriented
  - Java + Jena / TDB 0.8

Outline

Context

- Project GREASE-II

  Geographic Reasoning for Search Engines researches information access methods to large collections of documents and objects having geographically rich text and meta-data, with emphasis on the web.

- Required a boost in the development of the new version of its geospatial ontology of Portugal (Geo-Net-PT 02).
  - Add data of physical features.
  - Add geometric description of administrative features.

Extension of Geo-Net vocabulary

- Geo-Net-PT 02 requires explicit conceptualization of:
  - Name relationships (properties preferred, alternative, identifier)
  - Geographic codes (class GeographicCode, property inSchema)
  - Rich relationships between features (properties scope, isLocatedOn, is AdjacentTo, isConnectedTo, isPartOf, hasPart)
- Where does the extension stop?
  - When there is vocabulary available:
    - ownership of Web resources: SIOC vocabulary
    - owners of Web resources: FOAF vocabulary

Statoids

- Spatial extent: Portugal
- Temporal extent: Snapshot of summer 2009 of official available data
- Total concepts: 701,209
  - Features: 415,409
  - Names: 273,310
  - Complex footprints: 4,597
- Feature types: 81
- Sources: 21

Outline

Conclusions

- RDF "Little o" ontologies provide a flexible model and access interface to geographic ontologies. Experience shows:
  - Fast development of tools.
  - Uniform approach in different scenarios.
  - Clear separation of concerns.
    - Geo-Net deals with the structure of the world, but for describing the evolution of structure of the world use with a vocabulary for describing time, such as RDF Temporal .
- Geographic ontologies only show their usefulness if they are reused by third parties. Our next steps are:
  - Publish geographic terminologies as Linked Data.
  - Systematic Interlink with the Web of Data.

References

F. J. Lopez-Pellicer, M. Chaves, C. Rodrigues, and M. J. Silva, "Geographic Ontologies Production in GREASE-II," University of Lisbon, Faculty of Sciences, LaSIGE, Tech. Rep. TR 09-18, November 2009. [Online]. Available: http://hdl.handle.net/10455/3256

Geo-Net and Geo-Net-PT 02 are available through

`http://xldb.di.fc.ul.pt/wiki/Geo-Net-PT_02_in_English`

Thank you